The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification[1]

STEVEN K. HANKS* AND TONY HUNTER[2]

*Department of Cell Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA; and Molecular Biology and Virology Laboratory, The Salk Institute, San Diego, California 92186, USA

The eukaryotic protein kinases comprise one of the largest superfamilies of homologous proteins and genes. Within this family, there are now hundreds of different members whose sequences are known. Although there is a rich diversity of structures, regulation modes, and substrate specificities among the protein kinases, there are also common structural features. These conserved structural motifs provide clear indications as to how these enzymes manage to transfer the γ-phosphate of a purine nucleotide triphosphate to the hydroxyl groups of their protein substrates. The authors of this review have carried out a monumental task of analyzing and collating the amino acid sequences of all reported protein kinases and defining the conserved structural features that characterize the portion of these proteins that is responsible for their catalytic activity. Comparison of the sequences in the catalytic fragment of the protein kinases has been used to arrange these enzymes in evolutionary trees that group subfamilies of closely related enzymes. It is comforting that the structural relationships that emerge from these trees result in groupings that also reflect related functions. The work presented in this review seems to be an excellent example of the type of analysis that will become indispensable in the coming years, as more and more sequence information become available to biologists as a result of the genome projects.

ABSTRACT **The eukaryotic protein kinases make up a large superfamily of homologous proteins. They are related by virtue of their kinase domains (also known as catalytic domains), which consist of ~250–300 amino acid residues. The kinase domains that define this group of enzymes contain 12 conserved subdomains that fold into a common catalytic core structure, as revealed by the 3-dimensional structures of several protein-serine kinases. There are two main subdivisions within the superfamily: the protein-serine/threonine kinases and the protein-tyrosine kinases. A classification scheme can be founded on a kinase domain phylogeny, which reveals families of enzymes that have related substrate specificities and modes of regulation.—Hanks, S. K., Hunter, T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 9, 576–596 (1995)**

Key Words: protein-tyrosine kinase • protein-serine kinase • protein phosphorylation • AMP-dependent protein kinase

THE EUKARYOTIC PROTEIN KINASE SUPERFAMILY

One of the largest known protein superfamilies is made up of protein kinases identified largely from eukaryotic sources. (The term superfamily will be used here to distinguish this broad collection of enzymes from smaller, more closely related subsets that have been commonly referred to as families). These enzymes use the γ-phosphate of ATP (or GTP) to generate phosphate monoesters using protein alcohol groups (on Ser and Thr) and/or protein phenolic groups (on Tyr) as phosphate acceptors. The protein kinases are related by virtue of their homologous kinase domains (also known as catalytic domains), which consist of ~250–300 amino acid residues (reviewed in refs 1–3; and see below). During the past 15 years, previously unrecognized members of the eukaryotic protein kinase superfamily have been uncovered at an exponentially increasing rate and currently appear in the literature almost weekly. This pace of discovery can be attributed to the past development of molecular cloning and sequencing technologies and, more recently, to the advent of the polymerase chain reaction (PCR),[3] which facilitated the use of homology-based cloning strategies. Consequently, about 200 different superfamily members (products of distinct paralogous genes) had been recognized from mammalian sources alone! The prediction made several years ago (4) that the mammalian genome contains about 1000 protein kinase genes (roughly 1% of all genes) would still appear to be within reason, and may even be an underestimate (5).

In addition to mammals and other vertebrates, eukaryotic protein kinase superfamily members have been identified and characterized from a wide range of other animal phyla as well as from plants, fungi, and protozoans. Hence, the protein kinase progenitor gene can be traced back to a time before the evolutionary separation of the major eukaryotic kingdoms. The identification of eukaryotic-like protein kinase genes in prokaryotes (6, 7) raises the possibility that the protein kinase progenitor gene might have arisen before the divergence of prokaryotes and eukaryotes (see below). Studies of the budding and fission yeasts, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, have been particularly fruitful in the recognition of new protein kinases. In these geneti-

[1]This article is based on an introductory chapter in the *Protein Kinase Factsbook*, edited by D. G. Hardie and S. K. Hanks, published in 1995 by Academic Press, London.

[2]To whom correspondence and reprint requests should be addressed, at: Molecular Biology and Virology Laboratory, The Salk Institute, 10010 N. Torrey Pines Rd., La Jolla, CA 92037, USA.

[3]Abbreviations: PCR, polymerase chain reaction; PKA-Cα, type α cAMP-dependent protein kinase catalytic subunit; Cdk2, cyclin-dependent kinase 2; Erk2, p42 MAP kinase; APE,

0892-6638/95/0009-0576/$01.50.

cally tractable organisms, the powerful approach of mutant isolation and cloning by complementation has netted dozens of protein kinase genes required for numerous aspects of cell function (8). In many cases, vertebrate counterparts have now been found for these genes, leading to a growing awareness that protein phosphorylation pathways that regulate basic aspects of cell physiology have been maintained throughout the course of eukaryotic evolution.

Even though the overwhelming majority of protein kinases identified from eukaryotic sources belong to this superfamily, a small but growing number of such enzymes do not qualify as superfamily members. Most of these are related to the prokaryotic protein-histidine kinase family (see below), which forms the sensor components of two-component signal transduction systems (9). Included in this category are a putative ethylene receptor encoded by the flowering plant *ETR1* gene (10), the product of the budding yeast *SLN1* gene (11, 12) thought to be involved in relaying nutrient information to elements controlling cell growth and division, the mitochondrial branched-chain α-ketoacid dehydrogenase kinase (13), and the mitochondrial pyruvate dehydrogenase kinase (14). In prokaryotes, protein-histidine kinases phosphorylate aspartates in their target proteins, but except for the two dehydrogenase kinases that phosphorylate serine, the acceptor specificities of most of the eukaryotic protein kinases of this type are not known. In addition to these protein kinases, the Bcr protein encoded by the *breakpoint cluster region* gene involved in the Philadelphia chromosome translocation (15) and the A6 kinase isolated by expression cloning using an anti-phosphotyrosine antibody (16) have kinase domains unrelated to any known eukaryotic or prokaryotic kinase. In addition, true protein-histidine kinases are known in eukaryotes. One such enzyme has been extensively characterized from budding yeast but not yet molecularly cloned (17), and so it is not clear whether this enzyme will belong to the protein kinase superfamily or use a novel structural principle for phosphotransfer.

What about the prokaryotes? It has been known for years that protein phosphorylation events play key regulatory roles in numerous bacterial cell processes including chemotaxis, bacteriophage infection, nutrient uptake, and gene transcription (reviewed in refs 18, 19). The bacterial protein kinases have been divided into three general classes (20): *1)* protein-histidine kinases such as those functioning in two-component sensory regulatory systems (strictly speaking, these are protein-aspartyl kinases, because autophosphorylation on His is an intermediary step in phosphotransfer to an aspartate in the response-regulator protein) (9); *2)* phosphotransferases such as those of the phosphoenol pyruvate-dependent phosphotransferase system involved in sugar uptake (21); and *3)* protein-serine kinases such as isocitrate dehydrogenase kinase/phosphatase (22). Amino acid sequences have been determined for members of each class, and all are unrelated to the eukaryotic protein kinase superfamily.

Recently, however, true homologs of the eukaryotic protein kinases have been identified from two species of bacteria, *Yersinia pseudotuberculosis* (7) and *Myxococcus xanthus* (6, 23). Are these special cases, or the first examples of many such genes in prokaryotes? The eukaryotic-like protein kinase YpkA from the pathogenic enterobacteria *Y. pseudotuberculosis* is encoded by a plasmid essential for the virulence of this infectious organism. In addition to YpkA, at least two other proteins encoded by genes residing on the virulence plasmid exhibit high similarity to eukaryotic proteins. Thus, it seems likely that the virulence plasmid genes were transduced from a eukaryotic host by horizontal transfer. The myxobacterium *M. xanthus* presents a different and perhaps more intriguing picture. Application of the PCR homology-based cloning strategy revealed that at least eight genes encoding members of the eukaryotic protein kinase superfamily are present in the genome of this species (23). The myxobacteria are unusual prokaryotes in that they undergo a complex developmental cycle upon nutrient depletion, much like that of the eukaryotic slime mold *Dictyostelium*. Given that protein kinases are commonly involved in regulating growth and differentiation of eukaryotic cells, it is attractive to speculate that the eukaryotic-like protein kinases in *M. xanthus* are specifically involved in regulating their developmental cycle. Indeed, one of these kinases, Pkn1, was shown to be required for proper fruiting body formation. The same could be true for the eukaryotic-like protein kinase PknA from *Anabena* (24). In keeping with this idea, neither the PCR approach applied to *Escherichia coli* (23) nor extensive sequencing of the *E. coli* genome (now 30% complete) has yielded eukaryotic-like protein kinases. Hence, genes encoding members of the eukaryotic protein kinase superfamily may be present only in bacteria that can undergo a developmental cycle. However, unpublished reports of eukaryotic-like protein kinases in *Streptomyces coelicolor*, and in three species of *Methanococcus*, suggest that such genes are more widely expressed among prokaryotes, and potentially these genes represent the ancestors for the entire eukaryotic protein kinase superfamily.

THE HOMOLOGOUS KINASE DOMAINS

The kinase domains of eukaryotic protein kinases impart the catalytic activity. Three separate roles can be ascribed to the kinase domains: *1)* binding and orientation of the ATP (or GTP) phosphate donor as a complex with divalent cation (usually Mg^{2+} or Mn^{2+}); *2)* binding and orientation of the protein (or peptide) substrate; and *3)* transfer of the γ-phosphate from ATP (or GTP) to the acceptor hydroxyl residue (Ser, Thr, or Tyr) of the protein substrate.

Conserved features of primary structure

The total number of distinct kinase domain amino acid sequences available is now approaching 400 (**Table 1**). Included in this total are the vertebrate enzymes encoded by distinct paralogous genes, their presumed functional homologs from invertebrates and simpler organisms (encoded by orthologous genes), and those identified from lower organisms and plants for which vertebrate equivalents have not been found. Conserved features of kinase domain primary structure have previously been identified through an inspection of multiple amino acid sequence alignments (1-3) . The large number of sequences now available precludes showing an alignment containing all known kinase domains. Thus, in **Fig. 1** only 60 different kinase domain sequences are aligned. These are drawn, however, from the widest possible sampling of the superfamily and thus provide a good representation of the

Table 1. *Eukaryotic protein kinase superfamily classification.*

A-C-G Group

AGC-I. Cyclic nucleotide-regulated protein kinase family

A. Cyclic AMP-dependent protein kinase (PKA) subfamily

vertebrate:

1. PKA-Cα:	PKA catalytic subunit, alpha-form
2. PKA-Cβ:	PKA catalytic subunit, beta-form
3. PKA-Cγ:	PKA catalytic subunit, gamma-form

Drosophila melanogaster:

1. DmPKA-C0:	PKA catalytic subunit, C0 form
2. DmPKA-C1:	PKA catalytic subunit, C1 form
3. DmPKA-C2:	PKA catalytic subunit, C2 form

Caenorhabditis elegans:

1. CePKA:	PKA catalytic subunit homolog

Saccharomyces cerevisiae:

1. ScPKA-Tpk1:	PKA catalytic subunit homolog, type 1

Schizosaccharomyces pombe:

1. SpPKA1:	PKA catalytic subunit homolog

Dictyostelium discoideum:

1. DdPKA:	PKA catalytic subunit

Aplysia californica:

1. AplC:	PKA catalytic subunit homolog
2. Sak:	"Spermatozoon-associated kinase"

B. Cyclic GMP-dependent protein kinase (PKG) subfamily

vertebrate:

1. PKG-I:	PKG, type I
* 2. PKG-II:	PKG, type II

Drosophila melanogaster:

1. DmPKG-G1:	PKG homolog, type 1
2. DmPKG-G2:	PKG homolog, type 2

C. Others

Dictyostelium discoideum:

1. DdPK1:	PKA homolog

AGC-II. Diacylglycerol-activated/phospholipid-dependent protein kinase C (PKC) family

A. "Conventional" (Ca^{2+}-dependent) protein kinase C (cPKC) subfamily

vertebrate:

1. cPKCα:	Protein Kinase C, alpha-form
2. cPKCβ:	Protein Kinase C, beta-form
3. cPKCγ:	Protein Kinase C, gamma-form

Drosophila melanogaster:

1. DmPKC-53Ebr:	PKC homolog expressed in brain, locus 53E
2. DmPKC-53Eey:	PKC homolog expressed in eye, locus 53E

Aplysia californica:

1. Apl-I:	PKC homolog, type I

B. "Novel" (Ca^{2+}-independent) Protein Kinase C (nPKC) subfamily

vertebrate:

1. nPKCδ:	Protein Kinase C, delta-form
2. nPKCε:	Protein Kinase C, epsilon-form
3. nPKCη:	Protein Kinase C, eta-form
4. nPKCθ:	Protein Kinase C, theta-form

Drosophila melanogaster:

1. DmPKC-98F:	PKC homolog, locus 98F

Aplysia californica:

1. Apl-II:	PKC homolog, type II

Caenorhabditis elegans:

1. CePKC:	PKC homolog, product of *tpa-1* gene
* 2. CePKC1B:	PKC homolog expressed in neurons and interneurons

Dictyostelium discoideum:

* 1. DdMHCK:	PKC homolog

Saccharomyces cerevisiae:

1. ScPKA1:	PKC homolog, product of *PKC1* gene
* 2. ScPKA2:	PKC homolog, product of *PKC2* gene

Schizosaccharomyces pombe:

1. Pck1:	"Pombe C-kinase", type 1
2. Pck2:	"Pombe C-kinase", type 2

C. "Atypical" Protein Kinase C (aPKC) subfamily

vertebrate:

1. aPKCζ:	Protein Kinase C, zeta-form
* 2. aPKCι:	Protein Kinase C, iota-form
* 4. aPKCμ:	Protein Kinase C, mu-form

[a]More information about the individual protein kinases listed (including sequence references) can be obtained by contacting the authors or by consulting *The Protein Kinase Factsbook* (42). Protein kinases marked with asterisks (*) were not included in the phylogenetic analysis due to their recent discovery. In many instances new protein kinases were cloned by more than one group; in these cases the most commonly accepted name is used for the entry and alternative names are listed in parentheses after the entry. Protein kinase homologs from DNA viruses are not included in this classification.

Table 1. *(continued).*

D. Others
vertebrate:
* 1. PKN: Protein kinase with PKC-related catalytic domain

AGC-III. Related to PKA and PKC (RAC) family
vertebrate:
1. RAC-α: RAC, alpha-form; cellular homolog of v-Akt oncoprotein
2. RAC-β: RAC, beta-form
Drosophila:
1. DmRAC: RAC homolog
Caenorhabditis elegans:
* 1. CeRAC: RAC homolog

ACG-IV. Family of kinasese that phosphorylate G protein-coupled receptors
vertebrate:
1. βARK1: β-adrenergic receptor kinase, type 1
2. βARK2: β-adrenergic receptor kinase, type 2
3.RhK: Rhodopsin kinase
* 4.IT11: G-protein-coupled receptor kinase homolog
* 5.GRK5: G-protein-coupled receptor kinase, type 5
* 6. GRK6: G-protein-coupled receptor kinase, type 6
Drosophila melanogaster:
1. DmGPRK1: Drosophila G-protein-coupled receptor kinase, type 1
2. DmGPRK2: Drosophila G-protein-coupled receptor kinase, type 2

AGC-V. Family of budding yeast AGC-related kinases
Saccharomyces cerevisiae:
1. Sch9: Suppressor of defects in cAMP effector pathway
2. Ykr2: AGC-related kinase
3. Ypk1: AGC-related kinase

AGC-VI. Family of kinases that phosphorylate ribosomal S6 protein
vertebrate:
1. S6K: 70 kDa S6 kinase with single catalytic domain
2. RSK1(Nt): 90 kDA S6 kinase, type 1
3. RSK2(Nt): 90 kDA S6 kinase, type 2
[Note: The RSK enzymes have two distinct catalytic domains. The Nt-domain is closely related to S6K, whereas the Ct-domain is most closely related to phosphorylase kinase]

AGC-VII. Budding yeast Dbf2/20 Family
Saccharomyces cerevisiae:
1. Dbf2: Product of gene periodically expressed in cell cycle
2. Dbf20: Close relative of DBF2 not under cell cycle control

AG-VIII. Flowering plant "PVPK1 Family" of protein kinase homologs
Phylum Angiospermophyta (Kingdom Plantae):
1. PvK1: Bean protein kinase homolog
2. OsG11A: Rice protein kinase homolog
3. ZmPPK: Maize protein kinase homolog
4. AtPK5: Arabidopsis protein kinase homolog
5. AtPK7: Arabidopsis protein kinase homolog
6. AtPK64: Arabidopsis protein kinase homolog
7. PsPK5: Pea protein kinase homolog

Other AGC-related kinases
vertebrate:
1. DMPK: "Myotonic Dystrophy Protein Kinase"
2. Sgk: "Serum and glucocortocoid regulated kinase"
* 3. Mast205: Spermatid "Microtubule-associated serine/threonine kinase"
Neurospora crassa:
1. NcCot-1: Product of gene required for normal colonial growth
Dictyostelium discoideum:
1. Ddk2: Product of developmentally-regulated gene
Saccharomyces cerevisiae:
1. ScSpk1: Dual-specificity kinase
Phylum Angiospermophyta (Kingdom Plantae):
* 1. Atpk1: Arabidopsis protein kinase

CaMK Group
CaMK-I. Family of kinases regulated by Ca^{+}/Calmodulin, and close relatives
A. Subfamily including "Multifunctional" Ca^{+}/Calmodulin Kinases (CaMKs)
vertebrate:
1. CaMK1: CaMK, type I
2. CaMK2α: CaMK, type II, alpha subunit
3. CaMK2β: CaMK, type II, beta subunit
4. CaMK2γ: CaMK, type II, gamma subunit
5. CaMK2δ: CaMK, type II, delta subunit
* 6. EF2K: Elongation Factor-2 Kinase or CaMK type III
7. CaMK4: CaMK, type IV

Table 1. *(continued).*

Drosophila melanogaster:	
1. DmCaMK2:	CaMK-II homolog
Saccharomyces cerevisiae:	
1. ScCaMK2-1:	CaMK-II homolog, product of *CMK1* gene
2. ScCaMK2-2:	CaMK-II homolog, product of *CMK2* gene
Aspergillus nidulans:	
1. AnCaMK2:	CaMK-II homolog
B. Subfamily including phosphorylase kinases	
vertebrate:	
1. PhK-γM:	Skeletal muscle phosphorylase kinase catalytic subunit
2. PhK-γT:	Male germ cell phosphorylase kinase catalytic subunit
3. RSK1(Ct):	90 kDa S6 kinase, type 1; C-terminal catalytic domain
4. RSK2(Ct):	90 kDa S6 kinase, type 2; C-terminal catalytic domain
C. Subfamily including myosin light chain kinases	
vertebrate:	
1. skMLCK:	Skeletal muscle MLCK (rabbit)
2. smMLCK:	Smooth muscle MLCK (rabbit)
3. Titin:	Huge protein implicated in skeletal muscle development
Caenorhabditis elegans:	
1. Twn:	"Twitchin" protein involved in muscle contraction or development
Dictyostelium discoideum:	
1. DdMLCK:	Slime mold myosin light chain kinase
D. Subfamily of plant kinases with intrinsic calmodulin-like domain	
Phylum Angiospermophyta (Kingdom Plantae):	
1. CDPK:	Soybean Ca^{2+}-regulated kinase with intrinsic CaM-like domain
2. AtAK1:	Arabidopsis CDPK homolog
* 3. OsSpk:	Rice CDPK homolog
* 4. DcPk431:	Carrot CDPK homolog
E. Subfamily of plant kinases with highly acidic domain	
Phylum Angiospermophyta (Kingdom Plantae):	
* 1. ASK1:	Arabidopsis protein kinase homolog with highly acidic idomain
* 2. ASK2:	Arabidopsis protein kinase homolog with highly acidic domain
F. Other CaMK-related kinases	
vertebrate:	
1. PskH1:	Putative protein-serine kinase
* 2. MAPKAP2:	"MAP Kinase-Activated Protein Kinase 2"
Saccharomyces cerevisiae:	
1. Mre4:	Protein required for meiotic recombination
* 2. Dun1:	Protein required for DNA damage-inducible gene expression
* 3. Rck1:	"Radiation sensitivity complementing kinase, type 1"
* 4. Rck2:	"Radiation sensitivity complementing kinase, type 2"
CaMK-II. Snf1/AMPK family	
vertebrate:	
* 1: AMPK:	"AMP-Activated Protein Kinase"
2: p78:	Protein lost in carcinomas of human pancreas
Saccharomyces cerevisiae:	
1. Snf1:	Kinase essential for release from glucose repression
2. Kin1:	Protein kinase with N-terminal catalytic domain
3. Kin2:	Close relative of KIN1
4. Ycl24:	Protein kinase homolog on chromosome III
* 5. Ycl453:	Protein kinase homolog on chromosome XI
Schizosaccharomyces pombe:	
1. SpKin1:	Product of gene important for growth polarity
2. Nim1:	Inducer of mitosis
Phylum Angiospermophyta (Kingdom Plantae):	
1. PSnf1-RKIN1:	Rye putative protein kinase that complements yeast *snf1* polarity
2. PSnf1-AKIN10:	Arabidopsis putative protein kinase related to SNF1
3. PSnf1-BKIN12:	Barley protein related to SNF1
* 4. PKABA1:	Wheat kinase induced by abscisic acid
* 5. WPK4:	Wheat kinase homolog regulated by light and nutrients
* 6. NPK5:	Tobacco Snf1 homolog, activates *SUC2* gene expression
Other CaMK Group Kinases	
Plasmodium falciparum (malarial parasite):	
1. PfCPK:	Ca^{2+}-regulated kinase with intrinsic CaM-like domain
2. PfPK2:	Putative protein kinase
C-M-G-C Group	
CMGC-I. Family of cyclin-dependent kinases (CDKs) and other close relatives	
vertebrate:	
1. Cdc2:	Inducer of mitosis; functional homolog of yeast cdc2+/CDC28 kinases (Cdk1)
2. Cdk2:	Type 2 cyclin-dependent kinase
3. Cdk3:	Type 3 cyclin-dependent kinase
4. Cdk4:	Type 4 cyclin-dependent kinase
5. Cdk5:	Type 5 cyclin-dependent kinase

Table 1. *(continued).*

	6. Cdk6:	Type 6 cyclin-dependent kinase
	7. PCTAIRE1:	Cdc2-related protein
	8. PCTAIRE2:	Cdc2-related protein
	9. PCTAIRE3:	Cdc2-related protein
	10. Mo15:	"Cdk-activating kinase"; Negative regulator of meiosis (CAK)
	Drosophila melanogaster:	
	1. DmCdc2:	Functional homolog of yeast cdc2+/CDC28 kinases
	2. DmCdc2c:	Cdc2-cognate protein; Cdk2 homolog
	Dictyostelium discoideum:	
	1. DdCdc2:	Functional homolog of yeast cdc2+/CDC28 kinases
	2. DdPRK:	"Cdc2-related PCTAIRE Kinase"
	Aspergillus nidulans:	
	1. NIMXcdc2:	Cdc2-related gene product
	Plasmodium falciparum:	
	1. PfPK5:	Cdc2-related protein from human malarial parasite
	Entamoeba histolytica:	
	1. EhC2R:	Cdc2-related protein
	Crithidia fasciculata:	
	1. CfCdc2R:	Cdc2-related protein
	Leishmania mexicana:	
*	1. LmCRK1:	"Cdc2-Related Kinase"
	Saccharomyces cerevisiae:	
	1. Cdc28:	"Cell-division-cycle" gene product
	2. Pho85:	Negative regulator of the PHO system and cell cycle regulator
	3. Kin28:	CDC28-related protein
	Schizosaccharomyces pombe:	
	1. SpCdc2:	"Cell-division-cycle" gene product
	Histoplasma capsulatum:	
*	1. HcCdc2:	Cdc2 homolog from dimorphic fungus
	Phylum Angiospermophyta (Kingdom Plantae):	
	1. Pcdc2:	Flowering plant Cdc2 homolog othat complements yeast mutants
*	2. MsCdc2B:	Alfalfa Cdc2 cognate gene products that complements G1/S transition
	3. OsC2R:	More distantly related Cdc2 homolog from rice
CMGC-II. Erk(MAP kinase) family		
	vertebrate:	
	1. Erk1:	"Extracellular signal-regulated kinase", type 1 (p44 MAP kinase)
	2. Erk2:	"Extracellular signal-regulated kinase", type 2 (p42 MAP kinase)
	3. Erk3:	Somewhat distant relative of the Erk/MAP kinases
*	4. p63MAPK:	Another more distant relative of the Erk/MAP kinases
*	5. SAPK-α:	"Stress-activated protein kinase, type alpha" (JNK2)
*	6. SAPK-β:	"Stress-activated protein kinase, type beta"
*	7. SAPK-γ/Jnk1:	"Stress-activated protein kinase, type gamma" or "Jun N-terminal Kinase"
*	8. p38:	HOG1-related protein (MPK2)
	Drosophila melanogaster:	
	1. DmErkA:	Homolog of Erk/MAP kinases; product of *rolled* gene
	Caenorhabditis elegans:	
*	1. Sur1:	Erk/MAP kinase homolog
	Saccharomyces cerevisiae:	
	1. Kss1:	Suppressor of *sst2* mutant, overcomes growth arrest
	2. Fus3:	Product of gene required for growth and mating
	3. Slt2:	Product of gene complementing *lyt2* mutants (MPK1)
*	4. Hog1:	Product of gene required for osmoregulation
	Schizosaccharomyces pombe:	
	1. Spk1:	Product of gene that confers drug resistance to staurosporine, a PK inhibitor
	Phylum Deuteromycota (Kingdom Fungi):	
	1. CaErk1:	Protein that interferes with mating factor-induced cell cycle arrest
	Trypanosoma brucei (Phylum Zoomastigina, Kingdom Protoctista):	
*	1. KFR1:	"KSS1- and FUS3-related" gene product
	Phylum Angiospermophyta (Kingdom Plantae):	
	1. PErk:	Flowering plant Erk/MAP kinase homologs (7 distinct homologs identified in Arabidopsis)
CMGC-III. Glycogen synthase kinase 3 (GSK3) family		
	vertebrate:	
	1. GSK3α:	Glycogen synthase kinase 3, α-form
	2. GSK3β:	Glycogen synthase kinase 3, β-form
	Drosophila melanogaster:	
	1. Sgg:	Product of *shaggy/zeste-white 3* gene
	Saccharomyces cerevisiae:	
	1. Mck1:	"Meiosis and centromere regulatory kinase"
*	2. ScGSK3	Protein closely related to MCK1
*	3. Mds1:	Dosage suppressor of mck1 mutant
	Dictyostelium discoideum:	
*	1. DdGSK3:	Glycogen synthase kinase 3 homolog
	Phylum Angiospermophyta (Kingdom Plantae):	
*	1. ASK-α:	"Arabidopsis shaggy-related protein kinase", type alpha
*	2. ASK-γ:	"Arabidopsis shaggy-related protein kinase", type gamma

Table 1. *(continued).*

	vertebrate:	
	1. CK2α:	Casein kinase II, alpha subunit
	1. CK2α':	Casein kinase II, alpha-prime subunit
	Drosophila melanogaster:	
	1. DmCK2:	Casein kinase II homolog
	Caenorhabditis elegans:	
	1. CeCK2:	Casein kinase II homolog
	Theileria parva (a protozoan parasite):	
	1. TpCK2:	Casein kinase II α-subunit homolog
	Dictyostelium discoideum:	
	1. DdCK2:	Casein kinase II, α-subunit
	Saccharomyces cerevisiae:	
	1. ScCK2α:	Casein kinase II, alpha subunit
	2. ScCK2α':	Casein kinase II, alpha-prime subunit
	Schizosaccharomyces pombe:	
*	1. SpCka1:	Casein kinase II, α-subunit homolog (Orb5)
	Phylum Angiospermophyta (Kingdom Plantae):	
	1. ZmCK2:	Flowering plant casein kinase II, α-subunit homolog
	CMGC-IV. Clk family	
	vertebrate:	
	1. Clk:	"Cdc-like kinase"
*	2. Srpk1:	Kinase that regulates intracellular localization of splicing factors
	3. PskG1:	Putative protein kinase
	4. PskH2:	Putative protein kinase
	Drosophila melanogaster:	
*	1. Doa:	Kinase encoded by "Darkener of Apricot" locus
	Saccharomyces cerevisiae:	
	1. Yak1:	Suppressor of RAS mutant
	2. Kns1:	Nonessential protein kinase homolog
	Schizosaccharomyces pombe:	
	1. Dsk1:	Dis1-suppressing protein kinase implicated in mitotic control
*	2. Prp4:	Pre-mRNA processing gene product; lacks subdomains X-XI
	Other CMGC Group kinases	
	vertebrate:	
	1. Mak:	"Male germ cell-associated kinase"
	2. Ched:	"Cholinesterase-related cell division controller"
	3. PITSLRE:	Galactosyltransferase-associated kinase
	4. KKIALRE:	Cdc2-related protein
*	5. PITALRE:	Cdc2-related kinase
*	6. PISSLRE:	Cdc2-related kinase
	Saccharomyces cerevisiae:	
	1. Sme1:	Product of gene essential for start of meiosis
	2. Sgv1:	Kinase required for G-protein-mediated adaptive response to pheromone
	3. Ctk1:	Product of gene required for normal growth
	Phylum Angiospermophyta (Kingdom Plantae):	
*	1. Mhk:	Arabidopsis thaliana "Mak homologous kinase"
	Conventional Protein-Tyrosine Kinase Group (I-X: Non-membrane-spanning; XI-XXIII: Membrane-spanning)	
	PTK-I. Src family	
	vertebrate:	
	1. Src:	Cellular homolog of Rous sarcoma virus oncoprotein
	2. Yes:	Cellular homolog of Yamaguchi 73 sarcoma virus oncoprotein
	3. Yrk:	Yes-related kinase
	4. Fyn:	Protein related to Fgr and Yes
	5. Fgr:	Cellular homolog of Gardner-Rasheed sarcoma virus oncoprotein
	6. Lyn:	Protein related to Fgr and Yes
	7. Hck:	Hematopoietic cell protein-tyrosine kinase
	8. Lck:	Lymphoid T-cell protein-tyrosine kinase
	9. Blk:	Lymphoid B-cell protein-tyrosine kinase
*	10. Frk:	Fyn-related kinase
*	11. Rak:	STK-related kinase
*	12. Fyk:	"Fyn and Yes-related kinase" from electric ray
	Drosophila melanogaster:	
	1. DmSrc:	Src homolog, polytene locus 64B
	Dugesiai (Girardia) tigrina (Phylum Platyhelminthes):	
*	1. DtSpk-1:	"Src-like planarian kinase"
	Hydra vulgaris (Phylum Cnidaria):	
	1. Stk:	Src-related protein
	Spongilla lacustris (Phylum Porifera):	
	1. Srk1-4:	Four distinct Src-related kinases
	PTK-II. Brk family	
	vertebrate:	
*	1. Brk:	Protein-tyrosine kinase expressed in human breast tumors

Table 1. *(continued).*

PTK-III. Tec family	
vertebrate:	
1. Tec:	"Tyrosine kinase expressed in hepatocellular carcinoma"
2. Emt:	"Expressed mainly in T-cells" kinase (Itk, Tsk)
3. Btk:	"Bruton's agammaglobulinaemia tyrosine kinase" (Emb)
* 4. Txk:	Tec-related protein-tyrosine kinase
Drosophila melanogaster:	
1. DmTec:	Tec homolog, polytene locus 28C
PTK-IV. Csk family	
vertebrate:	
1. Csk:	"C terminal Src Kinase"; negative regulator of Src
* 2. MatK:	"Megakaryocyte-associated Tyr-kinase" (Hyl, Lsk, Ctk, Ntk)
PTK-V. Fes(Fps) family	
vertebrate:	
1. Fes/Fps:	Cellular homolog of feline and avian sarcoma viruses
2. Fer:	"Fes/Fps-related" kinase
Drosophila melanogaster:	
1. DmFer:	Fer-related protein
PTK-VI. Abl family	
vertebrate:	
1. Abl:	Cellular homolog of Abelson murine leukemia virus
2. Arg:	"Abl-related gene" product
Drosophila melanogaster:	
1. DmAbl:	Abl-related protein
Caenorhabditis elegans:	
1. CeAbl:	Nematode Abl-related protein
PTK-VII. Syk/Zap70 family	
vertebrate:	
1. Syk:	"Spleen tyrosine kinase"
2. Zap70:	T-cell receptor "zeta chain-associated protein of 70 kDa"
Hydra vulgaris (Phylum Cnidaria):	
* 1. Htk16:	Syk/Zap70-related
PTK-VIII. Jak family	
vertebrate:	
1. Tyk2:	Transducer of interferon α/β signals
2. Jak1:	"Janus kinase", type 1
3. Jak2:	"Janus kinase", type 2
* 4. Jak3:	"Janus kinase", type 3
Drosophila melanogaster:	
* 1. Hop:	Product of hopscotch gene required for establishing segmental body plan
PTK-IX. Ack	
vertebrate:	
* 1. Ack:	"CDC42Hs-associated kinase"
PTK-X. Fak	
vertebrate:	
1. Fak:	"Focal adhesion kinase"
PTK-XI. Epidermal growth factor receptor family	
vertebrate:	
1. EGFR:	Epidermal growth factor receptor
2. ErbB2:	Cell homolog of oncogene activated in ENU-induced rat neuroblastoma (Neu, HER2)
3. ErbB3:	Receptor tyrosine kinase related to EGFR (HER3)
4. ErbB4:	Receptor tyrosine kinase related to EGFR (Tyro2)
Drosophila melanogaster:	
1. DER:	Homolog of EGF receptor
Caenorhabditis elegans:	
1. LET-23:	Product of gene required for normal vulval development
Schistosoma mansoni (Phylum Platyhelminthes):	
1. SER:	EGF receptor homolog
PTK-XII. Eph/Elk/Eck receptor family	
vertebrate:	
1. Eph:	Kinase detected in "erythropoeitin-producing hepatoma"
2. Eck:	"Epithelial cell kinase"
3. Eek:	Eph/Elk-related protein-tyrosine kinase
4. Hek:	Eph/Elk related protein-tyrosine kinase (Cek4)
5. Sek:	"Segmentally-expressed kinase"
6. Elk:	"Eph-like kinase" detected in brain
* 7. Hek2:	"Human embryo kinase" type 2 (Cek10)
* 8. Htk:	"Hepatoma transmembrane kinase"
9. Cek5/Nuk:	"Chicken embryo kinase 5"/"Neural kinase"
* 10. Ehk1:	"Eph homology kinase-1" (Cek7)
* 11. Ehk2:	"Eph homology kinase-2"
* 12. Myk1:	"Mammary-derived tyrosine kinase, type 1"

Table 1. *(continued).*

*	13. Myk2:	"Mammary-derived tyrosine kinase, type 2"
*	14. Cek9:	"Chicken embryo kinase 9"
*	15. Pag:	"Pagliaccio" Xenopus protein expression in neural crest and neural tissues
*	16. Rtk1:	Zebrafish Eph/Elk-related protein-tyrosine kinase
*	17. Rtk2:	Zebrafish Eph/Elk-related protein-tyrosine kinase
*	18. Rtk3:	Zebrafish Eph/Elk-related protein-tyrosine kinase
PTK-XIII. Axl family		
vertebrate:		
	1. Axl:	"Anexelekto" (Gr. "uncontrolled") tyrosine kinase (UFO, Ark)
	2. Eyk:	Cellular homolog of RPL30 avian oncoprotein (c-Ryk)
*	3. Brt/Sky/Tif/Rse:	"Brain tyrosine kinase"/"Sea related protein tyrosine kinase"/"Tyrosine kinase with Ig-like and FN-III-like domains"/"Receptor sectaris" (Tyro3)
PTK-XIV. Tie/Tek family		
vertebrate:		
	1. Tie:	"Tyrosine kinase with Ig and EGF homology"
	2. Tek:	"Tunica interna endothelial cell kinase" (TIE2)
PTK-XV. Platelet-derived growth factor receptor family		
A. Subfamily witih 5 Ig-like extracellular domains		
vertebrate:		
	1. PDGFRα:	Platelet-derived growth factor receptor, type alpha
	2. PDGFRβ:	Platelet-derived growth factor receptor, type beta
	3. CSF1R:	Colony-stimulating factor-1 receptor (c-Fms)
	4. Kit:	Steel growth factor receptor
	5. Flk2:	"Fetal liver kinase-2" (Flt3)
B. Subfamily with 7 Ig-like extracellular domains		
vertebrate:		
	1. Flt1:	"Fms-like tyrosine kinase", type 1
	2. Flt4:	"Fms-like tyrosine kinase", type 4
	3. Flk1:	"Fetal liver kinase-1" (KDR)
PTK-XVI. Fibroblast growth factor receptor family		
vertebrate:		
	1. FGFR1:	Fibroblast growth factor receptor, type 1 (Flg, Cek1)
	2. FGFR2:	Fibroblast growth factor receptor, type 2 (Bek, K-SAM, Cek3)
	3. FGFR3:	Fibroblast growth factor receptor, type 3
	4. FGFR4:	Fibroblast growth factor receptor, type 4
Drosophila melanogaster:		
	1. DmFGFR1:	Fibroblast growth factor receptor homolog, type 1
*	2. DmFGFR2:	Fibroblast growth factor receptor homolog, type 2
PTK-XVII. Insulin receptor family		
vertebrate:		
	1. InsR:	Insulin receptor
	2. IGF1R:	Insulin-like growth factor receptor
	3. IRR:	Insulin receptor-related protein
Drosophila melanogaster:		
	1. DmInsR:	Homolog of insulin receptor
PTK-XVIII. Ltk/Alk family		
vertebrate:		
	1. Ltk:	"Leukocyte tyrosine kinase
*	2. Alk:	"Anaplastic lymphoma kinase
PTK-XIX. Ros/Sev family		
vertebrate:		
	1. Ros:	Cellular homolog of UR2 avian sarcoma virus oncoprotein
Drosophila melanogaster:		
	1. Sev:	Product of *sevenless* gene required for R7 photoreceptor cell development
PTK-XX. Trk/Ror family		
vertebrate:		
	1. Trk:	High molecular weight nerve growth factor receptor
	2. TrkB:	Receptor for nrain-derived neurotrophic factor and neurotrophin-4/5
	3. TrkC:	Trk-related protein; receptor for neurotrophin-3
	4. Ror1:	"Ror" putative receptor, type 1
	5. Ror2:	"Ror" putative receptor, type 2
	6. TcRTK:	Trk-related receptor (electric ray)
Drosophila melanogaster:		
*	1. Dror:	Putative neurotrophic receptor
PTK-XXI. Ddr/Tkt family		
*	1. Ddr:	"Discoidin Domain Receptor" (TrkE, CAK, NEP, Ptk3)
*	2. Tkt:	"Tyrosine Kinase Related to Trk" (Tyro 10)

Table 1. *(continued).*

PTK-XXII. Hepatocyte growth factor receptor family

vertebrate:

	1. HGFR:	Hepatocyte growth factor receptor (MET)
	2. Sea:	Cellular homolog of S13 avian erythroleukemia virus oncoprotein
	3. Ron:	"Recepteur d'Origine Nantaise"
*	4. Stk:	"Stem cell-derived tyrosine kinase"

PTK-XXIII. Nematode Kin15/16 family

Caenorhabditis elegans:

	1. CeKin15:	PTK expressed during hypodermal development
	2. CeKin16:	PTK expressed during hypodermal development

Other membrane-spanning protein-tyrosine kinases (each with no close relatives)

vertebrate:

	1. Ret:	Normal homolog of oncoprotein activated by recombination
	2. Klg:	"Kinase-like gene" product
*	3. Nyk/Ryk:	"Novel tyrosine kinase-related protein" (VIK, Mrk, Nbtk1)

Drosophila melanogaster:

	1. Torso:	Product of *torso* gene required for embryonic anterior/posterior determination
	2. DmTrk:	Distant relative of the mammalian trk gene

Marine sponge (Geodia cydonium):

*	1. GCTK:	Putative receptor PTK

Other protein kinase families (not falling into m ajor groups)

O-I. Polo family

vertebrate:

	1. Plk:	"Polo-like kinase"
	2. Snk:	"Serum-inducible kinase"
*	3. Sak:	Polo-related kinase isolated in screen for genes regulating sialylation

Drosophila melanogaster:

	1. Polo:	Protein kinase homolog required for mitosis

Saccharomyces cerevisiae:

	1. Cdc5:	Product of gene required for cell cycle progression

O-II. MEK/STE7 family

vertebrate:

	1. MEK1:	"MAP ERK Kinase", type 1
	2. MEK2:	"MAP ERK Kinase", type 2

Drosophila melanogaster:

	1. Dsor1:	

Saccharomyces cerevisiae:

	1. Ste7:	Kinase required for haploid-specific gene expression
	2. Pbs2:	Kinase required for antibiotic drug resistance
	3. Mkk1:	"MAP Kinase Kinase", type 1 (suppresses lysis defect of pkc1 mutant)
	4. Mkk2:	"MAP Kinase Kinase", type 2 (suppresses lysis defect of pkc1 mutant)

Schizosaccharomyces pombe:

	1. Byr1:	Kinase that suppresses ras1-mutant sporulation defect
	2. Wis1:	Suppressor of cdc phenotype in triple mutant *cdc25/wee1/win1* strains

O-III. MEKK/Ste11 family

vertebrate:

*	1. MEKK:	"MEK Kinase"

Saccharomyces cerevisiae:

	1. Ste11:	Protein required for cell-type-specific transcription
	2. Bck1:	"Bypass of C kinase" kinase

Schizosaccharomyces pombe:

	1. Byr2:	Product of gene required for pheromone signal transduction

Phylum Angiospermophyta (Kingdom Plantae):

*	1. NPK1:	Flowering plant (tobacco) homolog of Bck1

O-IV. Pak/Ste20 family

vertebrate:

*	1. Pak:	"p21-(Cdc42/Rac) activated kinase"

Saccharomyces cerevisiae:

	1. Ste20:	Product of gene required for pheromone response

O-V. NimA family

vertebrate:

	1. Nek1:	NimA-related kinase
*	2. Nek2:	NimA-related kinase (Nlk1)
*	3. Nek3:	NimA-related kinase
*	4. Nrk2:	NimA-related kinase
*	5. Stk1:	NimA-related kinase

Aspergillus nidulans:

	1. NIMA:	Cell cycle control protein kinase

Drosophila melanogaster:

	1. Fused:	Product of gene required for segment polarity

Table 1. *(continued).*

	Trypanosoma brucei (Phylum Zoomastigina, Kingdom Protoctista):	
	1. NrkA:	Trypanosome protein kinase related to NimA
	Saccharomyces cerevisaie:	
	1. Kin3:	Putative protein kinase
O-VI. wee1/mik1 family		
	vertebrate:	
	1. Wee1Hu:	Gene product able to complement S. pombe wee1 mutant
	Saccharomyces cerevisiae:	
*	1. Swe1:	Wee1 homolog from budding yeast
	Schizosaccharomyces pombe:	
	1. SpWee1:	"Wee" size at division kinase; Cdc2 negative regulator
	2. Mik1:	"Mitosis inhibitory kinase", negative regulator of Cdc2
O-VII. Family of kinases involved in translational control		
	vertebrate:	
	1. HRI:	"Heme-regulated eukaryotic initiation factor 2α kinase"
	2. PKR:	"Double-stranded RNA-dependent kinase" (Tik)
	Saccharomyces cerevisiae:	
	1. Gcn2:	Protein required for translational derepression
O-VIII. Raf family		
	vertebrate:	
	1. Raf-1:	Cellular homolog of retroviral oncogene product
	2. A-Raf:	Oncogenic protein closely related to c-Raf
	3. B-Raf:	Oncogenic protein closely related to c-Raf
	Drosophila melanogaster:	
	1. DmRaf:	Raf homolog
	Caenorhabditis elegans:	
	1. CeRaf:	Raf homolog; product of *lin-45* gene required for vulval differentiation
	Phylum Angiospermophyta (Kingdom Plantae):	
	1. Ctr1:	Negative regulator of ethylene response pathway
O-IX. Activin/TGFβ receptor family		
A. Subfamily of type I receptors		
	vertebrate:	
	1. ActR-I:	Type I receptor for activin and TGF-β (Tsk7L, SKR1, ALK-2)
*	2. TSR-1:	Type I receptor for activin and TGFG-β (ALK-1)
*	3. TGFβRI:	Type I receptor TGF- (ALK-5)
*	4. ActR-IB:	Type I receptor for activin (ALK-4)
*	5. BRK-1:	Type I receptor for BMP-2 and BMP-4 (ALK-3)
*	6. ALK-6:	"Activin receptor-like kinase", type 6
	Drosophila melanogaster:	
*	1. DmAtr-I:	Type I activin receptor homolog
*	2. DmSax:	Product of *saxophone* gene
B. Subfamily of type II receptors		
	vertebrate:	
	1. ActRII:	Type II receptor for activin
	2. ActRIIB:	Type II receptor for activin
	3. TGFβRII:	Type II receptor TGF-β
*	4. C14:	Putative receptor kinase expressed in gonads
	Drosophila melanogaster:	
*	1. DmAtr-II:	Type II activin receptor homolog
	Caenorhabditis elegans:	
*	1. DAF-4:	Larva development regulatory protein; BMP receptor
C. Others		
	Caenorhabditis elegans:	
	1. DAF-1:	Product of gene required for vulval development
O-X. Flowering plant putative receptor kinase family		
	Phylum Angiospermophyta (Kingdom Plantae):	
	1. ZmPK1:	Putative receptor protein-serine kinase (maize)
	2. Srk:	"S receptor kinase"; three distinct alleles: 2, 6, and 910 (Brassica)
	3. Tmk1:	Putative "Transmembrane receptor kinase" (Arabidopsis)
	4. Apk1:	Kinase that phosphorylates Tyr, Ser, and Thr (Arabidopsis)
*	5. Nak:	"Novel Arabidopsis Kinase" (Arabidopsis)
	6. Pro25:	Putative kinase selected for specificity to thylakoid membrane protein (Arabidopsis)
*	7. Pto:	Product of genen conferring pathogen resistance (tomato)
*	8. Tmkl1:	Transmembrane protein with unusual kinase-like domain (Arabidopsis)
*	9. Prk1:	Pollen-expressed receptor-like putative kinase (Petunia)
O-XI. Family of "mixed-lineage" kinases with leucine zipper domain		
	vertebrate:	
*	1. Mlk1:	"Mixed lineage kinase", type 1
*	2. Mlk2:	"Mixed lineage kinase", type 2
*	3. Mlk3:	"Mixed lineage kinase", type 3 (PTK1, SPRK)

Table 1. *(continued).*

O-XII. Casein kinase I family		
vertebrate:		
	1. CK1α:	Casein kinase I, type alpha
	2. CK1β:	Casein kinase I, type beta
	3. CK1γ:	Casein kinase I, type gamma
	4. CK1δ:	Casein kinase I, type delta
Saccharomyces cerevisiae:		
	1. Yck1:	Budding yeast casein kinase I homolog, type 1
	2. Yck2:	Budding yeast casein kinase I homolog, type 2
	3. Hrr25:	Kinase required for DNA repair
Schizosaccharomyces pombe:		
*	1. Hhp1:	Fission yeast casein kinase I homolog, type 1
*	2. Hhp2:	Fission yeast casein kinase I homolog, type 2
O-XIII. PKN family of prokaryotic protein kinases		
Myxococcus xanthus (Phylum Myxobacteria: Kingdom Prokaryotae):		
	1. Pkn1:	Protein kinase homologous to eukaryotic kinases
	2. Pkn2:	Protein kinase required for maintenance of stationary phase cells and development
Other protein kinase family members (each with no known close relatives)		
vertebrate:		
	1. Mos:	Cellular homolog of retroviral oncogene product
	2. Pim1:	Proto-oncogene activated by murine leukemia virus
	3. Cot:	Product of oncogene expressed in human thyroid carcinoma
	4. Esk:	"Embryonal carcinoma STY kinase"; dual specificity (PIT)
*	5. GC kinase:	Kinase expressed in germinal center B cells
*	6. Slk:	STE20-related kinase
*	7. LIMK:	"LIM motif-containing kinase"
*	8. Tsk1:	"Testis-specific kinase"
Drosophila melanogaster:		
	1. NinaC:	Product of gene essential for photoreceptor function
	2. Pelle:	Product of gene required for dorsalventral polarity
*	3. Nemo:	Product of gene required for rotation of photoreceptor clusters
Dictyostelium discoideum:		
	1. Spl A:	Spore lysis A protein kinase
	2. Dpyk2:	Developmentally-reguated tyrosine kinase, type 2
Ceratodon purpureus: (a moss)		
	1. PhyCer:	Putative protein-tyrosine kinase encoded by a phytochrome gene
Saccharomyces cerevisiae:		
	1. Cdc7:	"Cell-division-cycle" control gene product
	2. CDC15:	"Cell-division-cycle" control gene product
	3. Vps15:	Product of gene essential for sorting to lysosome-like vacuole
	4. Npr1:	Product of gene required for activity of ammonia-sensitive amino acid permeases
	5. Elm1:	Product of gene required for yeast-like cell morphology
	6. Ire1:	Required for Myo-inositol synthesis and signaling from ER to the nucleus
	7. Ykl516:	Putative protein kinase gene on chromosome XI
*	8. Ipl1:	Product of gene required for chromosome segregation
Schizosaccharomyces pombe:		
	1. Ran1:	Product of gene required for normal meiotic function
	2. Chk1:	"Checkpoint Kinase" that links rad pathway to Cdc2
*	3. Csk1:	"Cyclin Suppressing Kinase"
*	4. RPK1:	"Regulatory cell proliferation kinase"
Entamoeba histolytica (Phylum Rhizopoda, Kingdom Protoctista):		
	1. Ehmfk1:	Distant relative of Mos
Phylum Angiospermophyta (Kingdom Plantae):		
	1. GmPK6:	Protein kinase homolog (soybean)
*	2. Tsl:	Product of *Tousled* gene required for normal leaf/flower development (Arabidopsis)
Yersinia psuedotuberculosis (Phylum Omnibacteria, Kingdom Prokaryotae):		
	1. YpkA:	Enterobacterial protein kinase essential for virulence

known primary structures. The kinase domains are further divided into 12 smaller subdomains (indicated by Roman numerals), defined as regions never interrupted by large amino acid insertions and containing characteristic patterns of conserved residues (consensus line in Fig. 1).

Twelve kinase domain residues are recognized as being invariant or nearly invariant throughout the superfamily (conserved in over 95% of 370 sequences), and hence strongly implicated as playing essential roles in enzyme function. Using the type α cAMP-dependent protein kinase catalytic subunit (PKA-Ca) as a reference point, these are equivalent to Gly50 and Gly52 in subdomain I, Lys72 in subdomain II, Glu91 in subdomain III, Asp166 and Asn171 in subdomain VIB, Asp184 and Gly186 in subdomain VII, Glu208 in subdomain VIII, Asp220 and Gly225 in subdomain IX, and Arg280 in subdomain XI.

The patterns of amino acid residues found within subdomains VIB, VIII, and IX have been particularly well-conserved among the individual members of the dif-

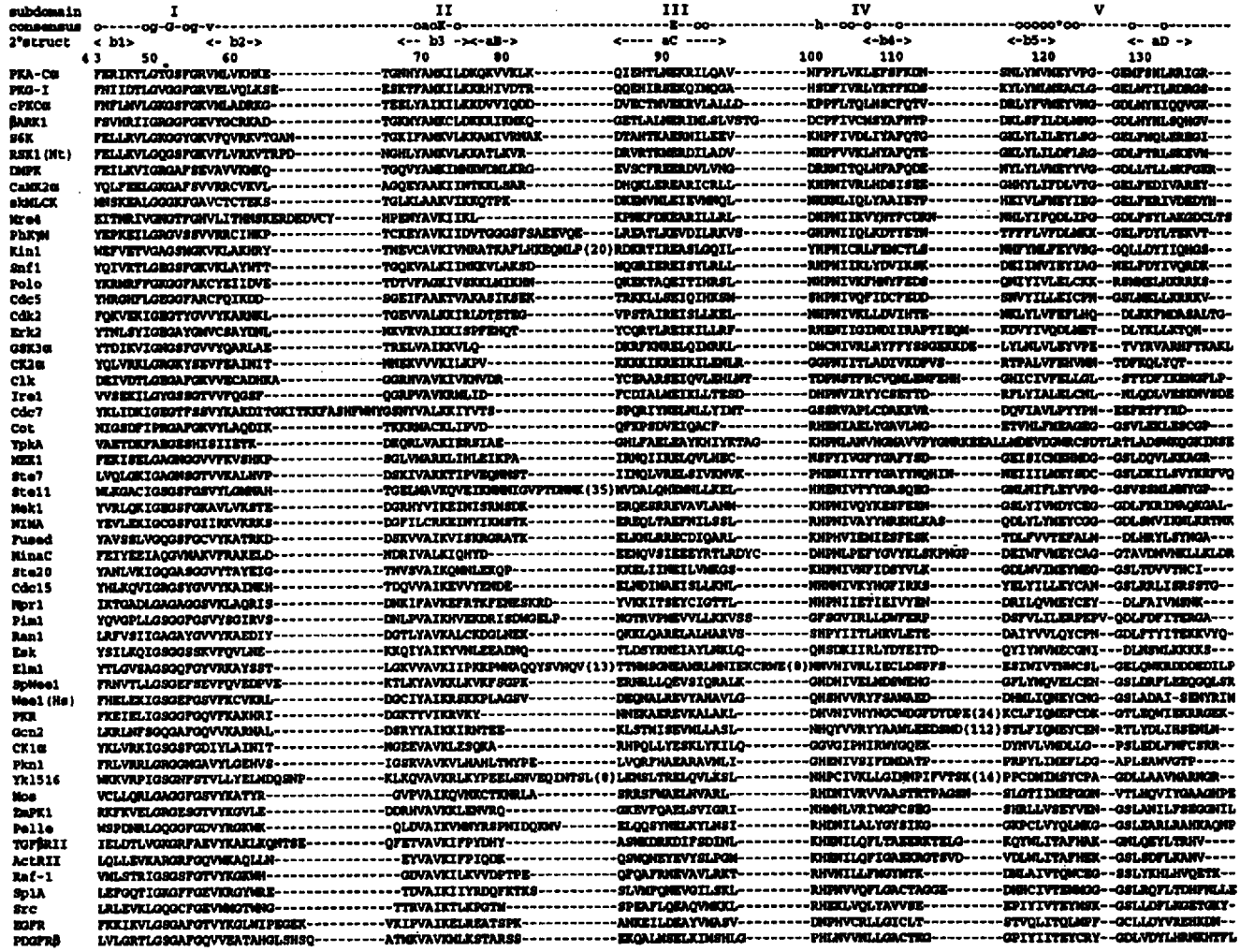


Figure 1. Multiple alignments of 60 kinase domains representative of members of the eukaryotic protein kinase superfamily. The abbreviated names used are as defined in Table 1. The single letter amino acid code is used and gaps are indicated by dashes. The entire sequences for the larger inserts are not shown, but excluded residues are indicated as numbers in brackets. Twelve distinct subdomains are indicated by Roman numerals. The consensus line is given according to the following code: uppercase letters, invariant residues, lowercase residues nearly invariant residues; o, positions conserving nonpolar residues; *, positions conserving polar residues; +, positions conserving small residues with near neutral polarity. Residues corresponding to the numbered β-strands (b) and α-helices (a) in PKA-Cα are indicated in the 2° structure line.

ferent protein kinase families and these motifs have been targeted most frequently in PCR-based homology cloning strategies aimed at identifying new family members.

Relationship between conserved subdomains, higher order structure, and catalytic mechanism

The homologous nature of the kinase domains implies that they all fold into topologically similar 3-dimensional core structures and impart phosphotransfer according to a common mechanism. The larger inserts found within some kinase domains are likely to represent surface elements that do not disrupt the basic core structure. With the solution of the crystal structure of mouse PKA-Cα, in a binary complex with a pseudosubstrate peptide inhibitor (PKI 5-24; TTYADFIASGRTGRRNAIHD, the underlined Ala substituting for the Ser phosphoacceptor), the general topology of a protein kinase catalytic core structure was revealed for the first time (25, 26). Later, structures of ternary complexes of PKA-Cα, the pseudosubstrate inhibitor, and either MgATP or MnAMP-PNP (an MgATP analog) were solved (27, 28). As a consequence of these studies, precise functional roles for most of the highly conserved kinase domain residues have now been assigned.

The kinase domain of PKA-Cα folds into a two-lobed structure (**Fig. 2**). The smaller, NH_2- terminal lobe, which includes subdomains I-IV, is primarily involved in anchoring and orienting the nucleotide. This lobe has a predominantly antiparallel β-sheet structure that is unique among nucleotide binding proteins. The larger COOH-terminal lobe, which includes subdomains VIA-XI, is largely responsible for binding the peptide substrate and initiating phosphotransfer. It is predominantly α-helical in content. Subdomain V residues span

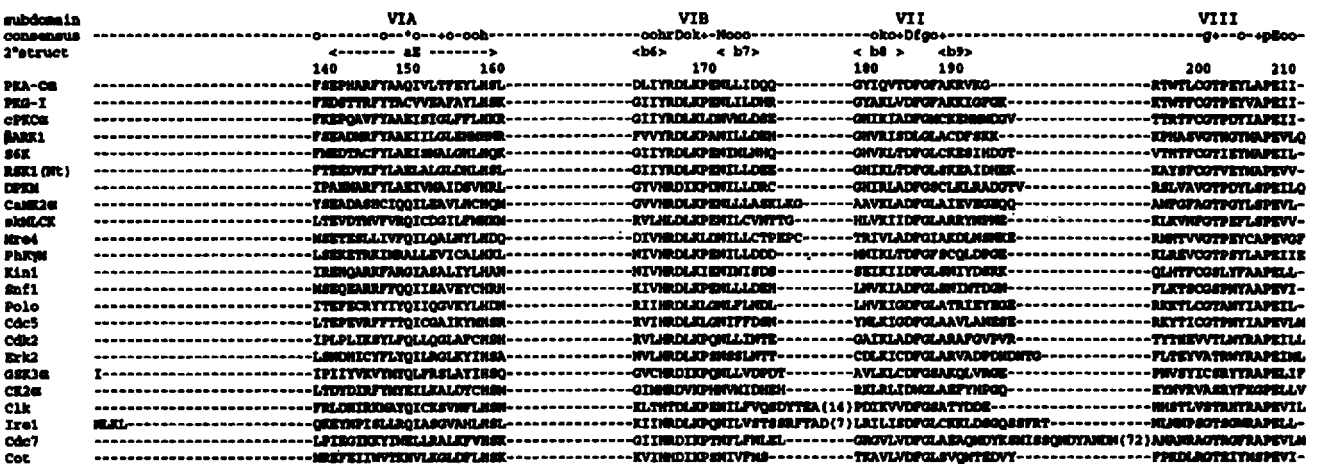


Figure 1 (contd.).

the two lobes. The deep cleft between the two lobes is recognized as the site of catalysis. The crystal structures of four additional eukaryotic protein kinase superfamily members—cyclin-dependent kinase 2 (Cdk2) (29), p42 MAP kinase (Erk2) (30), twitchin kinase (31), and casein kinase I (32)—have been reported more recently, and as expected, their kinase domains were found to fold into two-lobed structures topologically very similar to the catalytic core of PKA-Cα. Notable differences, however, were found in the regions corresponding to subdomain VIII in the Cdk2 and Erk2 structures, apparently reflecting the fact that these are structures of enzymes in an inactive state (see below). The twitchin structure is also of an inactive enzyme, but in this case it is inactive due to the presence of an autoinhibitory peptide sequence, which lies on the COOH-terminal side of the kinase domain and folds back into the active site cleft between the two lobes (31). This peptide apparently forces the two lobes to rotate almost 30° with respect to one another, and in this configuration inactive twitchin is more similar to the open configuration of PKA-Cα without PKI (33). In both twitchin and Cdk2 the α-helix C in subdomain III also adopts a different position to that of helix C in PKA-Cα. Unfortunately, no structure of a protein-tyrosine kinase catalytic domain was available at the time of writing (see "Note added in proof"), but the ease with which it has been possible to model the kinase domain of the EGF receptor protein-tyrosine kinase on to that of the PKA-Cα emphasizes that the structure of the protein-tyrosine kinases will be similar to that of the protein-serine kinases (34)

The conserved kinase subdomains correspond quite well to precise units of higher order structure. The functions of the individual subdomains will be discussed briefly later on a subdomain-by-subdomain basis, making reference to the crystal structure of PKA-Cα and

Figure 1 (contd.).

drawing attention to the proposed roles of the nearly invariant amino acid residues (25–27, 28) and other residues of interest. For more detailed information, the reader is referred to recent reviews on the structure of PKA-Cα (35–37) and to an excellent comparative review of the structures of PKA-Cα, Erk2, and Cdk2 (38).

Subdomain I, at the NH_2 terminus of the kinase domain, contains the consensus motif Gly-x-Gly-x-x-Gly-x-Val (starting with Gly50 in PKA-Cα). The kinase domain NH_2-terminal boundary occurs seven positions upstream of the first glycine in the consensus, where a hydrophobic residue is usually found. Subdomain I residues fold into a β-strand-turn-β-strand structure encompassing β-strands 1 and 2, and this structure acts as a flexible flap or clamp that covers and anchors the nontransferable phosphates of ATP. The backbone amides of Ser53, Phe54, and Gly55 form hydrogen bonds with ATP β- phosphate oxygens. Leu49 and Val57 contribute to a hydrophobic pocket that encloses the adenine ring of ATP.

Subdomain II contains the invariant Lys (Lys72 in PKA-Cα), which has long been recognized as being essential for maximal enzyme activity. This Lys lies within β-strand 3 of the small lobe, and helps anchor and orient ATP by interacting with the α- and β- phosphates. In addition, Lys72 forms a salt bridge with the carboxyl group of the nearly invariant Glu91 in subdomain III. Ala70 contributes to the hydrophobic adenine ring pocket. In PKA-Cα, β-strand 3 is followed immediately by α-helix B, which, judging from the sequence alignment, appears to be quite a variable structure among the protein kinases. Indeed, this α- helix is absent in the Cdk2 and Erk2 crystal structures.

Subdomain III represents the large α- helix C in the small lobe. The nearly invariant Glu residue (Glu91 in PKA-Cα) is centrally located in this helix and helps stabilize the interactions between Lys72 and the α- and β- phosphates of ATP. Subdomain IV corresponds to the hydrophobic β-strand 4 in the small lobe. This subdomain contains no invariant or nearly invariant residues

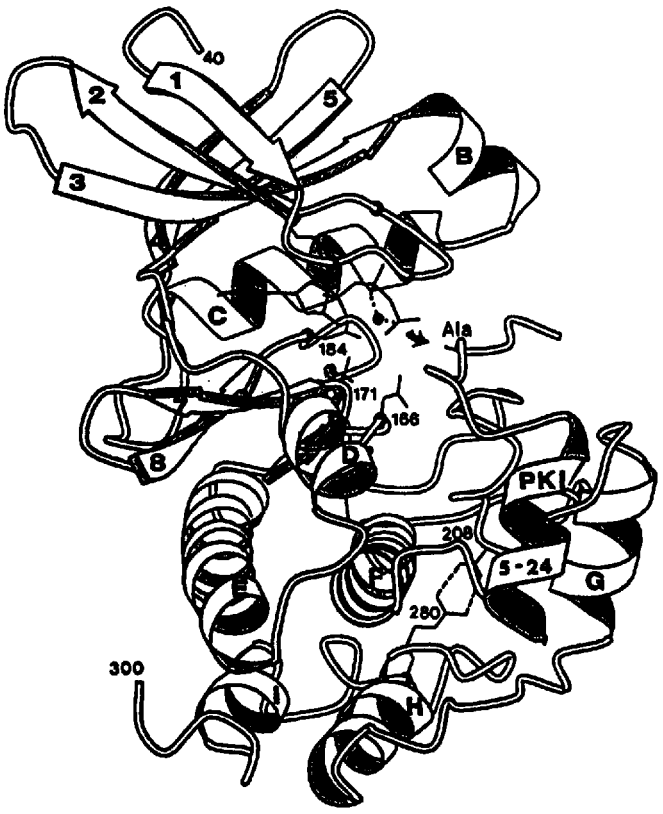


Figure 2. Ribbon diagram of the catalytic core of PKAα (residues 40–300) in a ternary complex with MgATP and pseudosubstrate peptide inhibitor (PKI -5–24). Invariant or nearly-invariant residues (Gly50, Gly52, Gly55, Lys72, Glu91, Asp166, Asn171, Asp184, Glu208, Asp220, and Arg280) are indicated by dots along the ribbon diagram. Side chains are shown for Lys72, Asp166, Asn171, Asp184, Glu208, and Arg280. β-strands and α-helices are indicated by flat arrow and helices, respectively, and are numbered according to Knighton et al. (26). The small arrow indicates the site of phosphotransfer with the Ala in PKI substituting for the phosphoacceptor Ser in the true substrate. (Reproduced, with permission, from Taylor et al. (36)).

and does not appear to be directly involved in catalysis or substrate recognition.

Subdomain V links the small and large lobes of the catalytic subunit and consists of the very hydrophobic β-strand 5 in the small lobe, the small α-helix D in the large lobe, and an extended chain that connects them. Three residues in the connecting chain of PKA-Cα, Glu121, Val123, and Glu127 help anchor ATP by forming hydrogen bonds with either the adenine or the ribose ring. Met120, Tyr122, and Val123 contribute to the hydrophobic pocket surrounding the adenine ring. Glu127 also participates in peptide binding by forming an ion pair with an Arg in the pseudosubstrate site of the PKA inhibitor peptide. This represents the first Arg in the PKA substrate recognition consensus Arg-Arg-x-Ser*-Hydrophobic.

Subdomain VIA folds into the large hydrophobic α-helix E that extends through the large lobe. None of the residues in helix E appear to interact directly with either MgATP or peptide substrate; hence this part of the molecule appears to act mainly as a support structure. Subdomain VIB folds into the small hydrophobic β-strands 6 and 7 with an intervening loop. Included here are two invariant residues (Asp166 and Asn171 in PKA-Cα) that lie within the consensus motif His-Arg-Asp-Leu-Lys-x-x-Asn (HRDLKxxN). The loop has been termed the catalytic loop because Asp166 within the loop has emerged as the likely candidate for the catalytic base, accepting the proton from the attacking substrate hydroxyl group during an in- line phosphotransfer mechanism. Lys168 in the loop (substituted by Arg in the conventional protein-tyrosine kinases) may help facilitate phosphotransfer by neutralizing the negative charge of the γ-phosphate during transfer. The side chain of Asn171 helps to stabilize the catalytic loop through hydrogen bonding to the backbone carbonyl of Asp166 and also acts to chelate the secondary Mg^{2+} ion that bridges the α- and γ-phosphates of the ATP. The carbonyl group of Glu170 forms a hydrogen bond with an ATP ribose hydroxyl group. Glu170 also participates in substrate binding by forming an ion pair with the second arginine of the peptide recognition consensus.

Subdomain VII folds into a β-strand-loop-b-strand structure, encompassing β-strands 8 and 9. The highly conserved DFG triplet, corresponding to Asp184-Phe185-Gly186 in PKA-Cα, lies in the loop that is stabilized by a hydrogen bond between Asp184 and Gly186. Asp184 chelates the primary activating Mg^{2+} ions that bridge the β- and γ-phosphates of the ATP, and thereby helps to orient the γ-phosphate for transfer. In Cdk2, β-strand 9 is replaced with a small α-helix designated αL12. However, it is unclear whether this helical character is maintained when Cdk2 is in its active conformation.

Subdomain VIII, which includes the highly conserved Ala-Pro-Glu ('APE') motif (residues 206–208 in PKA-Cα), folds into a tortuous chain that faces the cleft. Residues lying 7–10 positions immediately upstream of the APE motif are characteristically well-conserved among the members of different protein kinase families. The nearly invariant Glu corresponding to PKA-Cα Glu208 forms an ion pair with an invariant Arg (Arg280 in PKA-Cα) in subdomain XI, thereby helping to stabilize the large lobe.

Subdomain VIII appears to play a major role in recognition of peptide substrates. Several PKA-Cα subdomain VIII residues participate in binding the pseudosubstrate inhibitor peptide. Leu198, Cys199, Pro202, and Leu205 of PKA-Cα provide a hydrophobic pocket that accommodates the side chain of the hydrophobic residue at position +1 of the substrate consensus (Ile for the inhibitor peptide). Gly200 forms a hydrogen bond with the same Ile residue. Glu203 forms two ion pairs with the Arg in the high-affinity binding region of the inhibitor peptide.

Many protein kinases are known to be activated by phosphorylation of residues in subdomain VIII. In PKA-Cα, maximal kinase activity requires phosphorylation of Thr197, probably occurring through an intermolecular autophosphorylation mechanism (39). In the crystal structure, phosphate oxygens of phospho-Thr197 form hydrogen bonds with the charged side chains of Arg165, Lys189, and the hydroxyl group of Thr195, and thereby may act to stabilize the subdomain VIII loop in an active conformation permitting proper orientation of the substrate peptide. For members of the Erk (MAP) kinase family, phosphorylation of both a Thr and a Tyr

residue in subdomain VIII (mediated by members of the MEK kinase family) is required for activation. In the crystal structure determined for Erk2, these residues (Thr183 and Tyr185) were not phosphorylated and thus the enzyme was in an inactive state (unlike the PKA-Cα structure). The unphosphorylated Tyr185 is buried in a hydrophobic pocket, and interactions with Tyr185 are apparently required to hold the enzyme in the inactive state. Mutation of Tyr185, however, does not activate the enzyme, and so phosphorylation of Tyr185 must also play a role in activation. Unphosphorylated Erk2 appears to be inactive because residues required for catalysis are not properly oriented, and because its conformation results in a partial steric block to substrate binding. During activation of Erk2, Tyr185 phosphorylation precedes Thr183 phosphorylation; therefore, binding of MEK to Erk2 may alter the conformation of the subdomain VIII loop, thereby exposing Tyr185 for phosphorylation by MEK. Interaction of phospho-Tyr185 with surface residues would then allow the subdomain VIII loop to adopt the active conformation (30). Subsequent phosphorylation of the exposed Thr183 may activate the enzyme fully by promoting correct alignment of the catalytic residues. From the crystal structure of Cdk2, likewise in an inactive unphosphorylated state, the subdomain VIII loop appears to be in a conformation that would inhibit enzyme activity by sterically blocking the presumed protein substrate binding cleft (29). Phosphorylation of Thr160 in the Cdk2 subdomain VIII, mediated by MO15 (CAK), presumably would act to remove this inhibition by stabilizing the loop in an active conformation similar to that found in PKA-Cα. Cyclin binding to the NH_2-terminal lobe is also needed to activate Cdk2, and this may cause rotation of the NH_2-terminal domain resulting in correct alignment of catalytic residues.

Subdomain IX corresponds to the large α- helix F of the large lobe. The nearly invariant Asp corresponding to PKA-Cα Asp220 lies in the NH_2-terminal region of this helix and acts to stabilize the catalytic loop by hydrogen bonding to the backbone amides of Arg165 and Tyr164 that precede the loop. Glu230 of PKA-Cα forms an ion pair with the second Arg of the peptide recognition consensus. PKA-Cα residues 235-239 are all involved in hydrophobic interactions with the inhibitor peptide.

Subdomain X is the most poorly conserved subdomain and its function is obscure. In the crystal structure of PKA-Cα, it corresponds to the small α-helix G that occupies the base of the large lobe. Members of the Cdk, Erk (MAP), GSK3, and Clk kinase families (the C-M-G-C group) all have rather large insertions between subdomains X and XI, whose functional significance is presently unclear. Subdomain XI extends to the COOH-terminal end of the kinase domain. The most notable feature here is the nearly invariant Arg corresponding to Arg280 in PKA-Cα, which lies between α-helices H and I. The COOH-terminal boundary of the kinase domain is still poorly defined. For many protein-serine kinases, the consensus motif His-x-Aromatic-Hydrophobic is found beginning 9-13 residues downstream of the invariant Arg. For protein-tyrosine kinases, a hydrophobic amino acid lying 10 positions downstream of the invariant Arg appears to define the COOH-terminal boundary.

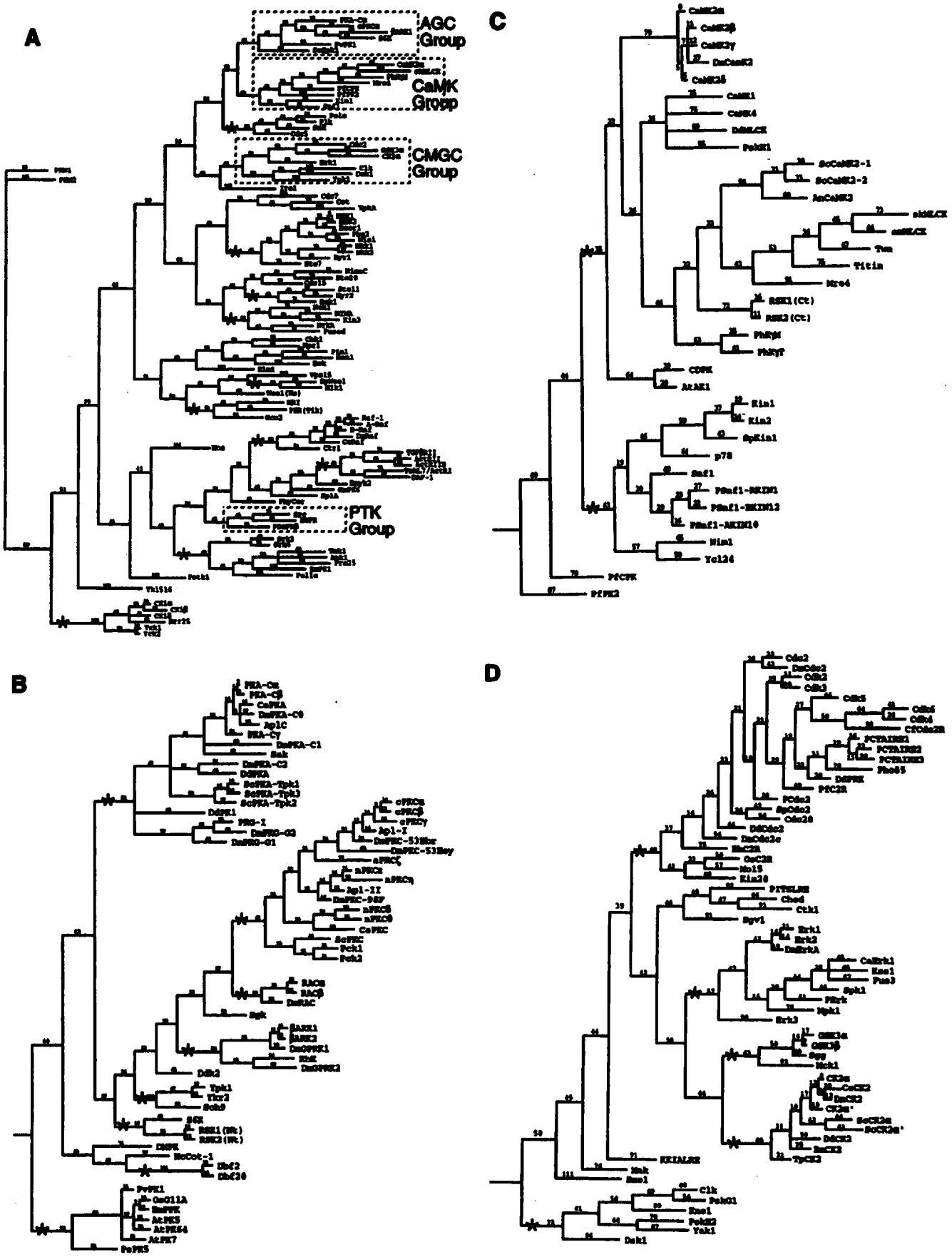
The amphipathic α-helix A of PKA-Ca (residues 15-35; not shown in Fig. 2), though lying outside of the conserved catalytic core on the NH_2-terminal side, appears to be an important feature found in many protein kinases (40). This helix spans the surface of both lobes of the core structure and complements and stabilizes the hydrophobic cleft between the two lobes. The A-helix motif appears to be present in many other protein kinases including members of the protein kinase C family and the Src family of protein-tyrosine kinases (40).

CLASSIFICATION OF EUKARYOTIC PROTEIN KINASES

To facilitate analysis and management of this large superfamily we have devised the classification scheme shown in Table 1, which subdivides the known members of the eukaryotic protein kinase superfamily into distinct families that share basic structural and functional properties. Phylogenetic trees derived from an alignment of kinase domain amino acid sequences (essentially an expanded version of Fig. 1) served as the basis for this classification. Thus, the sole consideration was similarity in kinase domain amino acid sequence. When considered alone, however, this property has been a good indicator of other characteristics held in common by the different members of the family.

Protein kinases whose entire kinase domain amino acid sequence had been published by July 1993 were included in phylogenetic analysis (as well as a few others made available at that time through sequence databases). If a given kinase domain sequence had been determined from more than one species among the vertebrates (i.e., orthologous gene products), only one representative (usually human) was included in the analysis. This policy was not used for the other phyla, however, because of greater divergences between the species and, hence, the sequences. The kinase domain phylogenies were inferred using the principle of maximum parsimony according to the PAUP software package developed by Swofford (41). Minimum-length trees were found using PAUP's 'heuristic' search method with branch swapping by the 'tree bisection-reconnection' strategy. Equal weights were given for all amino acid substitutions. Because multiple minimum-length trees were found, a consensus tree was calculated according to the method of Adams (cited in ref 41) in order to show branching ambiguities.

To accommodate the large numbers of sequences, it was necessary to construct five separate trees. Initially, a skeleton tree of 99 kinases was obtained (**Fig. 3*A***). The skeleton tree included only representative members from each of four large groups of protein kinases, each consisting of multiple related families known from previous work to cluster together in the tree. These four groups are designated: *1)* the AGC group, which includes the cyclic-nucleotide-dependent family (PKA and PKG), the protein kinase C (PKC) family, the β-adrenergic receptor kinase (βARK) family, the ribosomal S6 kinase family, and other close relatives; *2)* the CaMK group, which includes the family of protein kinases regulated by calcium/calmodulin, the Snf1/AMPK family, and other close relatives; *3)* the CMGC group, which includes the family of cyclin-dependent kinases, the Erk (MAP) kinase family, the glycogen synthase 3 (GSK3) family, the casein kinase II family, the Clk (Cdk-like kinase) family, and other close relatives; and *4)* the 'conventional' protein-tyrosine kinase (PTK) group. Separate trees (Fig. 3*B-E*) were later obtained for each of the four large kinase groups, and contain all members of the groups whose sequences were available at the time of analysis.


A
AGC Group
CaMK Group
CMGC Group
PTK Group
B
C
D

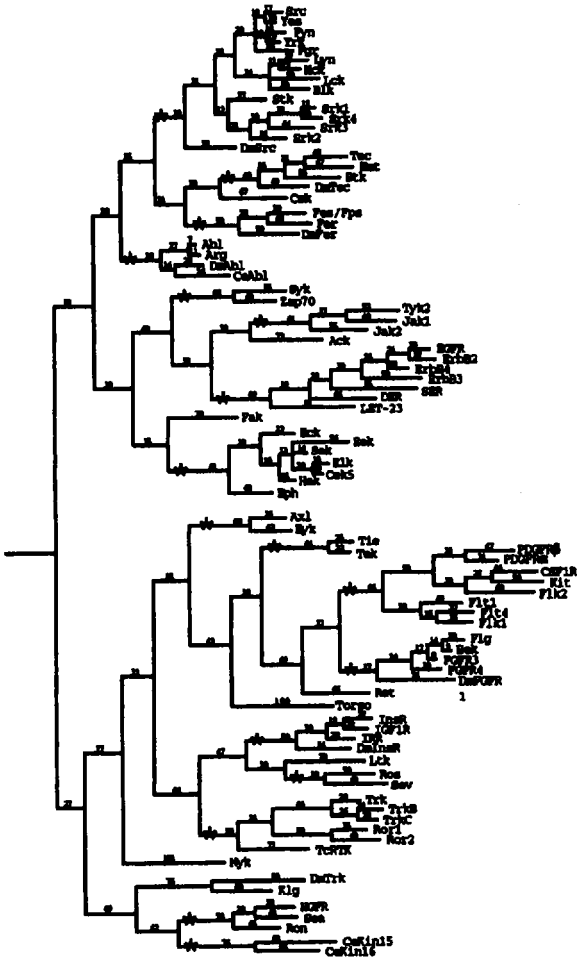


Figure 3. Phylogenetic trees of the eukaryotic protein kinase superfamily inferred from kinase domain amino acid sequence alignments. The abbreviated nomenclature is the same used in Table 1. *A)* 'Skeleton' tree showing 99 protein kinases. Positions of 4 clusters (AGC, CaMK, CMGC, and PTK) containing protein kinases representative of larger groups are indicated in the skeleton tree. *B)* AGC group tree of 59 protein kinases including PKA, PKG, and PKC and other close relatives. *C)* CaMK group tree of 35 protein kinases including the calcium/calmodulin-regulated enzymes. *D)* CMGC group tree of 59 protein kinases including the cyclin-dependent kinases. *E)* PTK group tree of 90 conventional protein-tyrosine kinases. Tree A is unrooted and drawn with Pkn1 and Pkn2 as outgroups. Outgroups of two or more distantly related protein kinases (not shown) were included in the analysis of trees B–E to provide a rooting point. Asterisks (*) in all trees indicate branches leading to defined protein kinase families listed in Table 1. Branch lengths indicate number of amino acid substitutions required to reach hypothetical common ancestors at internal nodes.

It can be reasonably surmised that the protein kinases having closely related catalytic domains, and thus defining a family, represent products of genes that have undergone relatively recent evolutionary separations. Given this, it should come as no surprise that members of a given family tend also to share related functions. This is manifest by similarities in overall structural topology, mode of regulation, and substrate specificity. The details of the common properties exhibited by the members of the various kinase families can best be gleaned from studying the information outlined in the individual entries section of the *Protein Kinase Factsbook* (42). Some of the most salient relationships are discussed below.

The AGC group protein kinases tend to be basic amino acid-directed enzymes, phosphorylating substrates at Ser/Thr residues lying very near Arg and Lys. For the cyclic nucleotide-dependent and ribosomal S6 kinase families, the preferred substrates have basic residues lying in specific positions NH_2-terminal to the phosphate acceptor. Preferred substrates for the PKC and RAC families have basic residues on both the NH_2- and COOH-terminal sides of the acceptor (43). The G-protein-coupled receptor kinases (βARK and RhK) appear to break this rule, however, as they are reported to prefer synthetic peptide substrate residues located within an acidic environment. Little substrate information is available for the other families in this group.

The CaMK group protein kinases also tend to be basic amino acid- directed, and in this regard it is notable that the AGC and CaMK groups fall near one another in the phylogenetic tree. CaMK1, CaMK2, CaMK4, MLCK, CDPK, and AMPK are all reported to prefer substrates with basic residues at specific positions NH_2-terminal to the acceptor site, whereas EF2K and PhK prefer sites with basic residues at both NH_2- and COOH-terminal locations. Many, but not all, of the CaMK group protein kinases are known to be activated by Ca^{2+}/calmodulin binding to a small domain located just COOH-terminal to the catalytic domain, e.g., CaMK1, CaMK2, CaMK4, PhKγ, MLCK, and twitchin. These enzymes and their close relatives are grouped together in a large family within the CaMK group. Also included in this family are a subfamily of plant enzymes (represented by CDPK) that contain an intrinsic calmodulin-like domain that confers Ca^{2+}-dependent activation. The other family within the CaMK group is the Snf1/AMPK family. Within this family, substrate specificity determinant information has been obtained only for the AMP-activated protein kinase, which also shows a requirement for an NH_2-terminal basic residue. The other major category of protein-serine kinases is the CMGC group. For the most part, these are proline-directed enzymes, phosphorylating substrates at sites lying in Pro-rich environments. Available data for Cdc2 and Cdk2 indicate that members of the cyclin-de-

pendent kinase family require phosphate acceptors lying immediately NH_2-terminal to a Pro. A similar requirement is indicated for the Erk (MAP) kinase family. The situation for the GSK3 family is more complicated, but most known acceptor sites lie within Pro-rich regions. The structures of Cdk2 and Erk2 indicate that the pocket for the +1 residue is shallower than in PKA-Cα due to the replacement of Leu205 by an Arg, which is bulkier and precludes binding of the larger hydrophobic amino acids. In addition, the unique secondary amide group of Pro may make special interactions (44). The casein-kinase II family enzymes fail to conform to the proline-directed specificity exhibited by the other major families of this group, showing instead a strong preference for Ser residues located NH_2-terminal to a cluster of acidic residues. The CMGC group protein kinases have larger-than-average kinase domains due to insertions between subdomains X and XI, whose functional significance is unknown.

The conventional protein-trosine kinase group includes a large number of enzymes with quite closely related kinase domains that specifically phosphorylate on Tyr residues (i.e., they cannot phosphorylate Ser or Thr). These enzymes, first recognized among retroviral oncoproteins, have been found only in metazoan cells where they are widely recognized for their roles in transducing growth and differentiation signals. Included in this group are more than a dozen distinct receptor families made up of membrane-spanning molecules that share similar overall structural topologies, and nine nonreceptor families also composed of structurally similar molecules. The specificity determinants surrounding the Tyr phosphoacceptor sites have yet to be firmly established for these enzymes, but Glu residues either on the NH_2- or COOH-terminal side of the acceptor are often preferred. This group is labeled "conventional" to distinguish it from other protein kinases (including Spk1, Clk, the MEK/Ste7 family members, Wee1/Mik1, ActRII, Hrr25, Esk, and Sp1A/DPyk2) reported to exhibit a dual specificity, that is, being capable of phosphorylating both Tyr and Ser/Thr residues (45). However, in most cases dual specificity has been observed only for autophosphorylation reactions in vitro, and the only dual specificity protein kinases that are known to be able to phosphorylate a substrate on Ser/Thr and Tyr are members of the MEK family. Considered as a group, these dual-specificity protein kinases are not particularly closely related to the conventional PTKs. Indeed, they seem to map throughout the phylogenetic tree (45), suggesting that the ability to autophosphorylate on Tyr may have had many independent origins during the evolutionary history of the superfamily.

The protein kinases falling outside the four major groups are a mixed bag. Although the individual members within the defined families found in this "other" category clearly are related to one another through both structure and function, it is difficult to make broader generalizations that could group any of these families together into a larger category. As far as substrate specificity determinants go, little is known about most "other" category protein kinases, due primarily to their rather recent discovery and the paucity of known physiological substrates. The casein kinase I family members, however, have been shown to prefer Ser/Thr residues located COOH-terminal to a phosphoserine or phosphothreonine, although a stretch of acidic residues may substitute. Also, the family of protein kinases involved in translational control (HRI, PKR/Tik, Gcn2) appear to be basic amino acid-directed enzymes preferring Ser residues lying NH_2- terminal to an Arg. Finally, as mentioned previously, the MEK/Ste7 family protein kinases and Wee1/Mik1 protein kinases exhibit a dual specificity.

Although this classification is based solely on catalytic domain sequences, members of families defined by this means are usually closely related in regions lying outside the cataytic domains and in many cases have been shown to possess very similar functions. Thus, intercalation of newly discovered protein kinases into this classification should allow one to make useful predictions about the functions of such enzymes.

FUTURE PROSPECTS

The rate of protein kinase discovery still shows no signs of abating. In addition to the continuing successes of homology-based approaches, genomic sequencing projects are beginning to make significant contributions. For instance, the sequences of two entire budding yeast chromosomes (46, 47) and a ~2 Mb stretch of *C. elegans* chromosome III (48) have revealed a number of new putative protein kinase genes. As genome sequencing projects gather speed, the number of new protein kinase genes discovered in this way will undoubtedly mushroom. This explosion of sequence data is making it increasingly difficult to manage protein kinase databases of the sort described here. Programs designed to align and derive relatedness trees are currently unable to handle the large number of available kinase domain sequences. New data handling programs will have to be developed to cope with large numbers of sequences like those of the eukaryotic protein kinase superfamily.

Protein kinase catalytic domain structures will continue to be solved. The first structure of a conventional protein-tyrosine kinase will be available shortly (see "Note added in proof"), and this should reveal how Tyr is selected as an acceptor amino acid vs. Ser/Thr. Such structures will enable comparative analysis to be carried out at the 3-dimensional level, and allow predictions of structures from primary sequences. Structural comparisons of catalytic domains with bound peptide substrates will also provide insights into substrate specificity. Most protein kinases show some degree of primary sequence specificity, and new methods are being developed to determine consensus sequence specificities for individual protein kinases (44). With such consensus information the structural basis for the binding of a preferred peptide sequence to the cognate substrate binding site can then be deduced. In the future, it may be possible to model the 3-dimensional structure of a novel protein kinase catalytic domain with sufficient accuracy to be able to deduce the preferred primary sequence surrounding the hydroxyamino acid it phosphorylates, which in turn will allow one to predict what proteins might be its substrates from the increasingly complete database of protein sequences. FJ

Note added in proof: The crystal structure of the tyrosine kinase domain of the insulin receptor has now appeared (Hubbard, S. R., Wei, L., Ellis, L., and Hendrickson, W. A. (1994) *Nature* **372**, 746-754).

REFERENCES

1. Hanks, S. K., Quinn, A. M., and Hunter, T. (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52
2. Hanks, S. K. (1991) Eukaryotic protein kinases. *Curr. Opin. Struct. Biol.* **1**, 369–3833.
3. Hanks, S. K., and Quinn, A. M. (1991) Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* **200**, 38–62
4. Hunter, T. (1987) A thousand and one protein kinases. *Cell* **50**, 823–8295.
5. Hunter, T. (1994) 1001 protein kinases redux: towards 2000. *Seminars Cell Biol.* In press
6. Muñoz-Dorado, J., Inouye, S., and Inouye, M. (1991) A gene encoding a protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium. *Cell* **67**, 995–1006
7. Galyov, E. E., Hakansson, S., Forsberg, A., and Wolf-Watz, H. (1993) A secreted protein kinase of *Yersinia pseudotuberculosis* is an indispensable virulence determinant. *Nature* **361**, 730–732
8. Hoekstra, M. F., DeMaggio, A. J., and Dhillon, N. (1991) Genetically identified protein kinases in yeast. Part 1: transcription, translation, transport and mating. *Trends Genet.* **7**, 256–261
9. Alex, L. A., Simon, M. J. (1994) Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. *Trends Genet.* **10**, 133–136
10. Chang, C., Kwok, S. F., Bleecker, A. B., and Meyerowitz, E. M. (1993) *Arabidopsis* ethylene-response gene *ETR1*: similarity of product to two-component regulators. *Science* **262**, 539–544
11. Ota, I. M., and Varshavsky, A. (1993) A yeast protein similar to bacterial two-component regulators. *Science* **262**, 566–569
12. Maeda, T., Wurgler-Murphy, S. M., and Saito, H. (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**, 242–245
13. Popov, K. M., Zhao, Y., Shimomura, Y., Kuntz, M. J., and Harris, R. A. (1992) Branched-chain α-ketoacid dehydrogenase kinase. *J. Biol. Chem.* **267**, 13127–13130
14. Popov, K. M., Kedishvili, N. Y., Zhao, Y., Shimomura, Y., Crabb, D. W., and Harris, R. A. (1993) Primary structure of pyruvate dehydrogenase kinase establishes a new family of eukaryotic protein kinases. *J. Biol. Chem.* **268**, 26602–26606
15. Maru, Y., and Witte, O. N. (1991) The BCR gene encodes a novel serine/threonine kinase activity within a single exon. *Cell* **67**, 459–468
16. Beeler, J. F., LaRochelle, W. J., Chedid, M., Tronick, S. R., and Aaronson, S. A. (1994) Prokaryotic expression cloning of a novel human tyrosine kinase. *Mol. Cell. Biol.* **14**, 982–988
17. Huang, J. M., Wei, Y. F., Kim, Y. H., Osterberg, L., and Matthews, H. R. (1991) Purification of a protein histidine kinase from the yeast *Saccharomyces cerevisiae*. The first member of this class of protein kinases. *J. Biol. Chem.* **266** 9023–9031
18. Stock, J. B., Ninfa, A. J., and Stock, A. M. (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol. Rev.* **53**, 450–490
19. Cozzone, A. J. (1993) ATP-dependent protein kinases in bacteria. *J. Cell. Biochem.* **51**, 7–13
20. Saier, M. H. (1993) Introduction: protein phosphorylation and signal transduction in bacteria. *J. Cell. Biochem.* **51**, 1–6
21. Reizer, J., Romano, A. H., and Deutscher, J. (1993) The role of phosphorylation of HPr, a phosphocarrier protein of the phosphotransfer system, in the regulation of carbon metabolism in gram-positive bacteria. *J. Cell. Biochem. 4751*, 19–24
22. LaPorte, D. C. (1993) Isocitrate dehydrogenase phosphorylation cycle: regulation and enzymology. *J. Cell. Biochem.* **51**, 14–18
23. Muñoz-Dorado, J., Inouye, S., and Inouye, M. (1993) Eukaryotic-like protein serine/threonine kinases in *Myxococcus xanthus*, a developmental bacterium exhibiting social behavior. *J. Cell. Biochem.* **51**, 29–33
24. Zhang, C.-C. (1993) A gene encoding a protein-related to eukaryotic protein kinases from the filamentous heterocystous cyanobacterium *Anabena* PCC7120. *Proc. Natl. Acad. Sci. USA* **90**, 11840–11844
25. Knighton, D. R., Zheng, J., Ten Eyck, L. F., Xuong, N.-H., Taylor, S. S., and Sowadski, J. M. (1991) Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 414–420
26 Knighton, D. R., Zheng, J., Ten Eyck, L. F., Ashford, V. A., Xuong, N.-H., Taylor, S. S., and Sowadski, J. M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407–420
27. Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H., and Huber, R. (1993) Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn^{2+} adenyl imidodiphosphate and inhibitor peptide PKI(5-24). *EMBO J.* **12**, 849–859
28. Zheng, J., Knighton, D. R., ten Eyck, L. F., Karlsson, R., Xuong, N., Taylor, S. S., and Sowadski, J. M. (1993) Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* **32**, 2154–2161
29. De Bondt, H. L., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgan, D. O., and Kim, S. (1993) Crystal structure of cyclin-dependent kinase 2. *Nature* **363**, 595–602
30. Zhang, F., Strand, A., Robbins, D., Cobb, M. H., and Goldsmith, E. J. (1994) Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. *Nature* **367**, 704–711
31. Hu, S.-H., Parker, M. W., Lei, J. Y., Wilce, M. C. J., Benian, G. M., and Kemp, B. E. (1994) Insights into autoregulation from the crystal structure of twitchin kinase. *Nature* **369**, 581–584
32. Carmel, G., Leichus, B., Cheng, X., Patterson, S. D., Mirza, U., Chait, B. T., and Kuret, J. (1994) Expression, purification, crystallization, and preliminary X-ray analysis of casein kinase-1 from *Schizosaccharomyces pombe*. *J. Biol. Chem.* **269**, 7304–7309
33. Zheng, J., Knighton, D. R., Xuong, N. H., Taylor, S. S., Sowadski, J. M., and Ten Eyck, L. F. (1993) Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Sci.* **2**, 1559nd573
34. Knighton, D. R., Cadena, D. L., Zheng, J., Ten Eyck, L. F., Taylor, S. S., Sowadski, J. M., and Gill, G. N. (1993) Structural features that specify tyrosine kinase activity deduced from homology modeling of the epidermal growth factor receptor. *Proc. Natl. Acad. Sci. USA* **90**, 5001–5005
35. Taylor, S. S., Knighton, D. R., Zheng, J., Ten Eyck, L. F., and Sowadski, J. M. (1992) Structural framework of the protein kinase family. *Annu. Rev. Cell Biol.* **8**, 429–462
36. Taylor, S. S., Zheng, J., Radzio-Andzelm, E., Knighton, D. R., Ten Eyck, L. F., Sowadski, J. M., Herberg, F. W., and Yonemoto, W. (1993) cAMP-dependent protein kinase defines a family of enzymes. *Phil. Trans. R. Soc. London B* **340**, 315–324
37. Madhusudan, A., Trafny, E. A., Xuong, N. H., Adams, J. A., Ten Eyck, L. F., Taylor, S. S., and Sowadski, J. M. (1994) cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer. *Protein Sci.* **3**, 176–187
38. Taylor, S. S., Radzio-Andzelm, E. (1994) Three protein kinase structures define a common motif. *Structure* **2**, 345–355
39. Steinberg, R. A., Cauthron, R. D., Symcox, M. M., and Shuntoh, H. (1993) Autoactivation of catalytic (Cα) subunit of cyclic AMP-dependent protein kinase by phosphorylation at threonine 197. *Mol. Cell. Biol.* **13**, 2332–2341
40. Veron, M., Radzio-Andzelm, E., Tsigelny, I., Ten Eyck, L. F., and Taylor, S. S. (1993) A conserved helix motif complements the protein kinase core. *Proc. Natl. Acad. Sci. USA* **90**, 10618–10622
41. Swofford, D. (1991) PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1. Ilinois Natural History Survey, Champaign, Illinois
42. Hardie, D. G., and Hanks, S. K. (1995) The Protein Kinase Factsbook, Academic Press, London
43. Pearson, R. B., Kemp, B. E. (1991) Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. *Meth. Enzymol.* **200**, 62–81
44. Songyang, Z., Blechner, S., Piwnica-Worms, H., and Cantley, L. C. (1994) A novel oriented peptide library technique for determining optimal substrates of protein kinases. *Curr. Biol.* In press
45. Lindberg, R. A., Quinn, A. M., and Hunter, T. (1992) Dual-specificity protein kinases: will any hydroxyl do? *Trends Biochem. Sci.* **17**, 114–119
46. Koonin, E. V., Bork, P., and Sander, C. (1994) Yeast chromosome III: new gene functions *EMBO J.* **13**, 493–503
47. Johnston, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Dover, J., Du, Z., Pavello, A., Fulton, L., Gattung, S., et al. (1994) Complete nucleotide sequence of Saccharomyces cerevisiae chromosome VIII. *Science* **265**, 2077–2082
48. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J ., Burton, J., Connell, M., Copsey, T., Cooper, J., et al. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.. *Nature* **368**, 32–38